Progress Report on AISR Grant NNG06GE71G
Principal Investigator: István Szapudi
Scalable Algorithms for Fast Analysis of Megapixel CMB Maps
and Large Astronomical Databases
Institute for Astronomy, Hawaii

Up to December 2007

We have introduced new extensions into our software pipeline based on SpICE that enables calculation of the cumulant correlator power spectrum. We demonstrated that it works in a real world environment by by calculating the 2-1 cumulant correlator power spectrum and WMAP3. A new method for calculating the full bispectrum based on our previous results on three-point functions is under way. We started to develop a new method, based on shrinkage estimators, which has a potential to improve statistical estimates in many areas of astrophysics. We first applied this idea for determining the covariance matrix of power spectrum measurement, and we are working on applying the same idea for C_l estimation of the CMB.

1 Principal Results

1.1 SpICE and WMAP

We have been continuing to work on our successful software package SpICE and the WMAP specific pipeline based on it. In Chen & Szapudi (2006) we demonstrated the new capabilities by measuring the 2-1 cumulant correlator power spectrum C_l^{21}, a degenerate bispectrum, from the second data release of the Wilkinson Microwave Anisotropy Probe (WMAP). Our high resolution measurements with SpICE span a large configuration space ($\simeq 168 \times 999$) corresponding to the possible cross-correlations of the maps recorded by the different differencing assemblies. We developed a novel method to recover the eigenmodes of the correspondingly large Monte Carlo covariance matrix. We examined its eigenvalue spectrum and used random matrix theory to show that the off diagonal terms are dominated by noise. The care and difficulty we needed to apply for calculating a useful covariance matrix motivated our project on shrinkage estimators (see below). We minimize the χ^2 to obtain constraints for the non-linear coupling parameter $f_{NL} = 22 \pm 52(1\sigma)$.

1.2 Three-point statistics

While the 2-1 cumulant correlator captures a lot of information about the bispectrum, we have been working on calculating the *full* bispectrum based on the three-point function code we developed at the end of the previous AISR. We have extended the numerical methods developed in (Szapudi et al., 2005, also supported by the previous AISR) to turn the problem into a double Hankel-integration. This involved turning these ideas to code by developing a robust numerical Hankel-transformation package using a `python` framework based on the `numpy` and `scipy`. We are on the verge of obtaining the first full bispectrum measurement using this package. Parts of this software package is integrated in our recently developed `cosmopy` framework, a preview (ver 0.3) of which is experimentally available from our group homepage `http://www.ifa.hawaii.edu/cosmowave`.

1.3 Shrinkage Estimators

The experience during the Monte Carlo estimation of the covariance matrix for the 2-1 cumulant correlator power spectrum motivated us to take a hard look at MC covariance matrix estimation: this is a problem which permeates all statistical analysis of the CMB and large scale structure as well. While the focus has been so far on calculating an accurate power spectrum (C_l's), less attention has been focused on the covariance matrix, which is equally important. We have performed full scale Monte Carlo simulations to gauge how errors propagate from the covariance matrix to parameter estimation and we were surprised by the degree to which an otherwise good measurement can yield misleading results due to tiny errors in the covariance matrix.

To remedy the situation, in Pope & Szapudi (2007) we introduced a novel statistical technique, shrinkage estimation, to estimate the power spectrum covariance matrix from a limited number of simulations. Although the idea has been around since the late 50's in the statistics literature, only recent results made it algorithmically feasible to a range of astrophysical problems, in particular covariance matrix estimation. To our knowledge we are the first to apply this technique to astrophysics.

We optimally combined an empirical estimate of the covariance with a model (the target) to minimize the total mean squared error compared to the true underlying covariance. We test our technique on N-body simulations and evaluate its performance by estimating cosmological parameters. Using a simple diagonal target, we show that the shrinkage estimator significantly outperforms both the empirical covariance and the target individually when using a small number of simulations. We find that reducing noise in the covariance estimate is essential for properly estimating the values of cosmological parameters as well as their confidence intervals. We extend our method to the jackknife covariance estimator and again find significant improvement, though simulations give better results. Even for thousands of simulations we still find evidence that our method improves estimation of the covariance matrix. Because our method is simple, requires negligible additional numerical effort, and produces superior results we always advocate shrinkage estimation for the covariance of the power spectrum and other large-scale structure measurements when purely theoretical modeling of the covariance is insufficient. These ideas, although first demonstrated on LSS power spectrum, should be directly applicable to the CMB power spectrum as well.

2 Acknowledged Support

A list of publications acknowledging AISR support is the following:

References

Chen, G., & Szapudi, I. 2006, ApJLett, 647, L87

Neyrinck, M. C., & Szapudi, I. 2007a, ArXiv e-prints, 710

Neyrinck, M. C., & Szapudi, I. 2007b, MNRAS, 375, L51

Pan, J., Coles, P., & Szapudi, I. 2007, MNRAS, 1066

Pope, A. C., & Szapudi, I. 2007, ArXiv e-prints, 711

Szapudi, I., Pan, J., Prunet, S., & Budavári, T. 2005, ApJLett, 631, L1